

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



(43) Date de la publication internationale
27 octobre 2005 (27.10.2005)

PCT

(10) Numéro de publication internationale
WO 2005/101292 A3

(51) Classification internationale des brevets⁷ : **G06K 9/00**

(21) Numéro de la demande internationale :
PCT/FR2005/000673

(22) Date de dépôt international : 18 mars 2005 (18.03.2005)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
0403556 5 avril 2004 (05.04.2004) FR

(71) Déposant et

(72) Inventeur : **LEBRAT, François** [FR/FR]; 98 avenue de
Versailles, F-75016 Paris (FR).

(74) Mandataires : **HASSINE, Albert** etc.; Cabinet Plasser-
aud, 65/67, rue de la Victoire, F-75440 Paris Cedex 09
(FR).

(81) États désignés (sauf indication contraire, pour tout titre de
protection nationale disponible) : AE, AG, AL, AM, AT,
AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO,
CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB,
GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP,
KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK,
MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL,
PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM,
ZW.

(84) États désignés (sauf indication contraire, pour tout titre
de protection régionale disponible) : ARIPO (BW, GH,

[Suite sur la page suivante]

(54) Title: METHOD FOR SEARCHING CONTENT PARTICULARLY FOR EXTRACTS COMMON TO TWO COMPUTER
FILES

(54) Titre : PROCÉDÉ DE RECHERCHE DE CONTENU, NOTAMMENT D'EXTRAITS COMMUNS ENTRE DEUX
FICHIERS INFORMATIQUES

AA			L	a	t	o	r	t	u	e
AD1	SN1		0	1	2	3	4			
BB			0	?	1	1	0			
L	0	0								
e										
	1	0								
i										
è	2	?								
v										
r	3	1								
e										
	4	0								

AA...THE TORTOISE

BB...THE HARE

retaining the other address pairs that identify data packets that may comprise said common extract.

(57) Abstract: The present invention relates to searching content, particularly for at least one extract common to a first data file and a second data file. The method comprises a preliminary step of preparing at least the first file by (a) dividing the first file into a series of data packets having a predetermined size, and identifying packet addresses in said file, (b) combining each packet address with a digital signature that defines one of three fuzzy logic states, namely true, false and indeterminate, and is the result of a combinatorial computation on data from said file; thereafter said method comprises performing an actual search for a common extract by (c) comparing the fuzzy logic states combined with each packet address of the first file with fuzzy logic states determined on the basis of data from the second file, and (d) removing from said common extract search the respective address pairs from the first and second files that have the respective logic states true and false or false and true, and

[Suite sur la page suivante]

WO 2005/101292 A3



GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Déclaration en vertu de la règle 4.17 :

- *relative à la qualité d'inventeur (règle 4.17.iv)) pour US seulement*

Publiée :

- *avec rapport de recherche internationale*

- *avant l'expiration du délai prévu pour la modification des revendications, sera republiée si des modifications sont reçues*

(88) Date de publication du rapport de recherche internationale:

26 janvier 2006

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

(57) Abrégé : L'invention concerne la recherche de contenu, notamment une recherche d'au moins un extrait commun entre un premier fichier et un second fichier de données. Le procédé de l'invention comporte une préparation préalable du premier fichier au moins, comprenant les étapes suivantes : a) segmenter le premier fichier en une succession de paquets de données, de taille choisie, et identifier des adresses de paquets dans ledit fichier, b) associer à l'adresse de chaque paquet une signature numérique définissant un état en logique floue parmi au moins trois états : "vrai", "faux" et "indéterminé", ladite signature résultant d'un calcul combinatoire sur des données issues dudit fichier, et le procédé se poursuit par une recherche d'extrait commun, proprement dite, comprenant les étapes suivantes, c) comparer les états de logique floue associés à chaque adresse de paquet du premier fichier, avec des états de logique floue déterminés à partir de données issues du second fichier, d) éliminer de ladite recherche d'extrait commun des couples d'adresses respectives des premier et second fichiers dont les états logiques respectifs sont "vrai" et "faux" ou "faux" et "vrai", et conserver les autres couples d'adresses identifiant des paquets de données susceptibles de comporter ledit extrait commun.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/FR2005/000673

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 G06K9/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G06K G06F H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, INSPEC, IBM-TDB, PAJ, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T	<p>SNASEL V: "Fuzzy signatures for multimedia databases"</p> <p>ADVANCES IN INFORMATION SYSTEMS. FIRST INTERNATIONAL CONFERENCE, ADVIS 2000. PROCEEDINGS (LECTURE NOTES IN COMPUTER SCIENCE VOL.1909) SPRINGER VERLAG BERLIN, GERMANY, 2000, pages 257-264, XP008038840 ISBN: 3-540-41184-4</p> <p>p. 259-260, paragraphe "4. fuzzy Signatures"</p> <p>-----</p> <p>-/--</p>	1-30

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

7 November 2005

Date of mailing of the international search report

14/11/2005

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Granger, B

INTERNATIONAL SEARCH REPORT

International Application No
PCT/FR2005/000673

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T	S. HABER ET AL: "If Piracy is the Problem, Is DRM the Answer?" DIGITAL RIGHTS MANAGEMENT. TECHNOLOGICAL, ECONOMIC, LEGAL AND POLITICAL ASPECTS, 31 August 2003 (2003-08-31), pages 224-233, XP002307417 HEIDELBERG p. 229, paragraphe "III.4 Fuzzy Hashing"	1-30
T	FALOUTSOS C: "ACCESS METHODS FOR TEXT" ACM COMPUTING SURVEYS, NEW YORK, NY, US, vol. 17, no. 1, March 1985 (1985-03), pages 49-74, XP000747611 ISSN: 0360-0300 p.54-56, paragraphe "2.3 Multiattribute Retrieval Methods"	1-30

RAPPORT DE RECHERCHE INTERNATIONALE

Demande Internationale No
PCT/FR2005/000673

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
CIB 7 G06K9/00

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)

CIB 7 G06K G06F H04L

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)

EPO-Internal, INSPEC, IBM-TDB, PAJ, COMPENDEX

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie °	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
T	<p>SNASEL V: "Fuzzy signatures for multimedia databases"</p> <p>ADVANCES IN INFORMATION SYSTEMS. FIRST INTERNATIONAL CONFERENCE, ADVIS 2000. PROCEEDINGS (LECTURE NOTES IN COMPUTER SCIENCE VOL.1909) SPRINGER VERLAG BERLIN, GERMANY, 2000, pages 257-264, XP008038840</p> <p>ISBN: 3-540-41184-4</p> <p>p. 259-260, paragraphe "4. fuzzy Signatures"</p> <p>-----</p> <p>-/--</p>	1-30



Voir la suite du cadre C pour la fin de la liste des documents



Les documents de familles de brevets sont indiqués en annexe

° Catégories spéciales de documents cités:

"A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent

"E" document antérieur, mais publié à la date de dépôt international ou après cette date

"L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)

"O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens

"P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

"&" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

7 novembre 2005

Date d'expédition du présent rapport de recherche internationale

14/11/2005

Nom et adresse postale de l'administration chargée de la recherche internationale

Office Européen des Brevets, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Granger, B

RAPPORT DE RECHERCHE INTERNATIONALE

Demande Internationale No
PCT/FR2005/000673

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
T	<p>S. HABER ET AL: "If Piracy is the Problem, Is DRM the Answer?" DIGITAL RIGHTS MANAGEMENT. TECHNOLOGICAL, ECONOMIC, LEGAL AND POLITICAL ASPECTS, 31 août 2003 (2003-08-31), pages 224-233, XP002307417 HEIDELBERG p. 229, paragraphe "III.4 Fuzzy Hashing"</p> <p>-----</p>	1-30
T	<p>FALOUTSOS C: "ACCESS METHODS FOR TEXT" ACM COMPUTING SURVEYS, NEW YORK, NY, US, vol. 17, no. 1, mars 1985 (1985-03), pages 49-74, XP000747611 ISSN: 0360-0300 p.54-56, paragraphe "2.3 Multiattribute Retrieval Methods"</p> <p>-----</p>	1-30